STOCKHOLM SCHOOL OF ECONOMICS
MASTER'S THESIS IN FINANCE

Robustness of Conditional Value-at-Risk (CVaR) for Measuring Market Risk

Mattias Letmark[a] & Markus Ringström[b]

[a] 18691@student.hhs.se; [b] 18416@student.hhs.se

Tutor: Assistant Professor Joel Reneby
Presentation date: June 16, 2006 13.15-15.00
Venue: Stockholm School of Economics, room 550

Acknowledgements

We would like to thank our tutor, Joel Reneby, for providing helpful feedback, support and valuable advice in writing this thesis.

Abstract

In this thesis the risk measure Conditional Value-at-Risk (CVaR) is studied in terms of robustness and whether it is an unbiased measure. The scope of the study is market risk. The results indicate that it is possible to construct an unbiased and robust CVaR measure in most cases, but that it is important to be careful when choosing the parameters of the CVaR estimator. However, in some cases CVaR does not seem to be unbiased or robust, primarily when applied to individual stocks.

Contents

1 Introduction

This section provides an introduction to the subject and an overview of previous work. Furthermore, we present the contribution and purpose of the thesis.

1.1 Background

1.1.1 Market Risk Management

Increasing trading activities and large portfolios held by participants on financial markets have made the measurement of market risk a primary concern for regulators and risk managers. Coordinated by the Basel Committee on Banking Supervision, banks are required to hold a certain amount of capital against adverse market movements. Specifically, banks must demonstrate that its capital is sufficient to cover losses 99.9% of the times over a one year holding period [11]. Such a risk capital is usually called Value-at-Risk (VaR).

An important milestone in the development of VaR models was JPMorgan's decision in 1994 to make its VaR system, RiskMetrics [39], available on the Internet. In the following years the RiskMetrics system essentially attained the status of a de-facto standard within the financial industry and a benchmark for measuring market risk. However, in the financial literature, additional measures of market risk besides VaR have been studied. Artzner et al. [8] highlighted some theoretical shortcomings of VaR as a measure of market risk. For example, it does not take into account the magnitude of losses when VaR is exceeded. VaR also fails to meet the characteristic of subadditivity (see section 2.2.2), i.e. the risk of a portfolio in terms of VaR may be larger than the sum of its components. Artzner et al. [8] proposed an alternative risk measure defined as the expected value of losses exceeding the VaR. This new risk measure has sounder theoretical properties (e.g. fulfills the subadditivity condition) and is usually called Conditional Value-at-Risk (CVaR).

1.1.2 Asset Allocation and Portfolio Theory

Asset allocation is always a topicality and Markowitz's portfolio theory has influenced academia and financial institutions since it was published in 1952 (see [33]). Markowitz proposed that a portfolio should be optimized in a mean-variance framework, i.e. maximizing the returns and at the same time keeping the risk under control. The definition of risk in this framework was defined as the overall portfolio variance. A more comprehensive description of portfolio theory and portfolio optimization is given in [14].

A drawback related to variance as a risk measure is that it penalizes upside (gains) and downside (losses) equally. As a complement to the mean-variance optimization model, not only relying on the variance as a risk measure, additional constraints can be added to control the risk. This is especially important as a tool for agency control. Alexander and Baptista [5] analyze the results from imposing VaR and CVaR constraints in the mean-variance framework. They show that in some cases such impositions may induce perverse effects, e.g. that risk averse agents select portfolios with larger standard deviations.

Instead of optimizing according to the mean-variance model, a portfolio can be optimized in other frameworks. Since VaR is one of the most popular risk measures in risk management, many studies have been performed on optimization in the mean-VaR framework. However, a problem that arises is that the optimization process is very complex, e.g. much more complex than optimization in the mean-CVaR framework. Uryasev and Rockafellar [42] proposed a mean-CVaR model using a linear optimization method and showed that VaR was calculated as a by-product. Another advantage with the mean-CVaR model is that CVaR optimization is more stable over different confidence levels, at least in the case of fixed-income securities (see [34]). Olszewski [37] studied hedge funds and suggested that a more efficient portfolio can be constructed by optimization in the mean-CVaR domain compared to the classic mean-variance domain.

1.1.3 Contribution

A conclusion so far is that there exist a large number of different risk measures, of which only a few have been mentioned here, all with its own characteristics, advantages and flaws. VaR has been adopted as the main measure of market risk, triggered by regulatory authorities such as the Basel Committee on Banking Supervision (Basel [44]) and also by authorities of the EU member states, e.g. the Swedish Financial Supervisory Authority (Finansinspektionen [45]).

However, VaR has attracted a lot of criticism as a risk measure. One reason is that the VaR concept can lead to perverse effects if used as a control mechanism. An example is shown in [15], where it is described how to earn $1 million in just one week with no initial capital. The catch is that there is a risk, even though it is highly improbable, of losing a *huge* amount of money. Other drawbacks with VaR are e.g. that using VaR as a risk measure may fail to stimulate diversification, due to its non-subadditivity characteristic (see e.g. [40]), and that VaR only provides a point-estimate of the loss distribution. The VaR estimate does not provide any information on the losses in the tail exceeding VaR, i.e. information on so-called "spike the firm" events (low probability, high loss) is not captured with the model. Yet recent history has shown that such events pose a real threat to e.g. the banking system (see [18]).

In the financial literature it is often suggested that CVaR is a better risk measure and has sounder properties than VaR. But is CVaR really better than VaR? CVaR would at least solve some problems, e.g. the diversification problem stated in [40] (due to its subadditivity property) and the situation where you could earn $1M with no initial capital described in [15] (due to its consideration of high, low-probability losses). Furthermore, in [19] it is shown that different assets will be ranked in the same way in terms of risks measured as VaR and CVaR, respectively. At least, this indicates that the good properties of VaR in some sense is transferred to the properties of CVaR and that CVaR in terms of risk ranking seems to do an equally good job as VaR.

There has been extensive research on CVaR in terms of portfolio optimization, but to our knowledge not much in terms of robustness. This topic will be further investigated in the thesis. A similar study was also suggested as future work in [29].

1.2 Purpose

The purpose of this thesis is to study the robustness of CVaR as a risk measure and also to study whether it is an unbiased estimate, and if so, under what circumstances it is robust and unbiased. In addition, we will compare different asset classes in terms of CVaR robustness and bias.

1.3 Outline

The structure of the thesis is as follows. Section 1 provides an introduction to the subject and an overview of previous work. Section 2 introduces the theoretical framework. We define the two risk measures VaR and CVaR and discuss the concept of robustness. In section 3 we present how we select the data sample used in the thesis. In section 4 we develop two hypotheses about the robustness of CVaR. Section 5 describes the methodology and the empirical measures used in the analysis. In section 6 we present the results of different tests to examine the robustness of CVaR. Eventually, section 7 concludes the main results and provides suggestions for further research.

2 Theoretical Framework

This section introduces the theoretical framework for measuring market risk. We define the two risk measures VaR and CVaR and discuss some of their properties. We also give a brief overview of two major estimation techniques, delta-normal approach and historical simulation. Moreover, we describe backtesting procedures and eventually discuss the concept of robustness.

2.1 Market Risk

Participants of financial markets face a risk of disastrous losses due to unexpected adverse movements in market factors. The risk of losses arising from movements in market prices is often referred to as *market risk*. The Basel Committee on Banking Supervision [10] classifies the sources of market risk into four main categories: equities, interest rate related instruments, foreign exchange and commodities. Over the last years, we have seen an increasing instability in the financial environment, an increasing globalization of financial markets, a significant growth of trading activity, development of numerous new financial products, new enabling technologies and regulatory requirements. These are all factors contributing to an increasing interest in market risk.

There are two main approaches of measuring market risk, *statistical methods* and *scenario based methods*. Comprehensive risk managers combine the use of statistical risk measures with techniques such as stress testing, scenario analysis and visualization. Just as a single diagnostic such as body temperature is not a reliable measure of the health of a human being, risk managers should not rely solely on a single method to determine the health (risk) of a portfolio.

2.2 Risk Measures

In this thesis we focus on statistical risk measures. Since the pioneering work of Markowitz [33], where he introduced the modern portfolio theory, the variance and standard deviation have been the traditional risk measures in economics and finance. However, there are several shortcomings related to these risk measures. For example, the variance penalizes upside (gains) and downside (losses) equally and mean-variance decisions are usually not consistent with the expected utility approach, unless returns are normally distributed or a quadratic utility function is used. Moreover, variance does not account for fat tails of the underlying distribution and therefore is inappropriate to describe the risk of low probability events, such as default risks.

2.2.1 Value-at-Risk

In recent years, academics and practitioners have extensively studied a risk measure called Value-at-Risk (VaR). It was developed to respond to the need to aggregate the various sources of market risk into a single quantitative measure. VaR focuses on the downside risk of a portfolio and is defined as the maximum expected loss at a specific confidence level (e.g. 95%) over a certain time horizon[1] (e.g. ten days). For example, if VaR is -$100 for a portfolio at a confidence level of 95% and a time horizon of one week we can state that "with 95% certainty we will not lose more than $100 over the next week". In another example, consider a bank that calculates its VaR assuming a one-day holding period and a 99% confidence level. Then the bank can expect that, on average, trading losses will exceed the VaR on one occasion in one hundred trading days.

The choice of confidence level varies among different risk managers. For example, the Basel Committee recommends the 99.9% confidence level for capital adequacy purposes [11]. For internal use, lower confidence levels is often used. For example, JPMorgan [27] uses the 99% level, Citibank [16] uses a confidence level of 95.4% and Goldman Sachs [24] uses the 95% level.

Another parameter that varies among risk managers is the time horizon (holding period) over which VaR is estimated. It is likely that the portfolio return changes more over a month than over a single day. The length of the holding period depends on the nature of the portfolio and typically ranges from one day to one month. The Basel Committee recommends a time horizon of ten days for most capital market transactions [11].

The mathematical definition of VaR is:

$$\int_{-\infty}^{\mathrm{VaR}_\alpha} f_X(x)\mathrm{d}x = 1-\alpha \tag{1}$$

or equivalently

$$P[x \leq \mathrm{VaR}_\alpha] = 1-\alpha \tag{2}$$

where $f(x)$ is the marginal probability function of portfolio returns x over the given time period and the confidence level is $\alpha \in [0,1]$.

A graphical interpretation of VaR using a confidence level of 95% is illustrated in Figure 1. VaR is the cut-off point separating the return distribution from its 5% tail.

[1] Since VaR assumes no changes in the portfolio weights during the time horizon, the term *holding period* is often used instead of time horizon

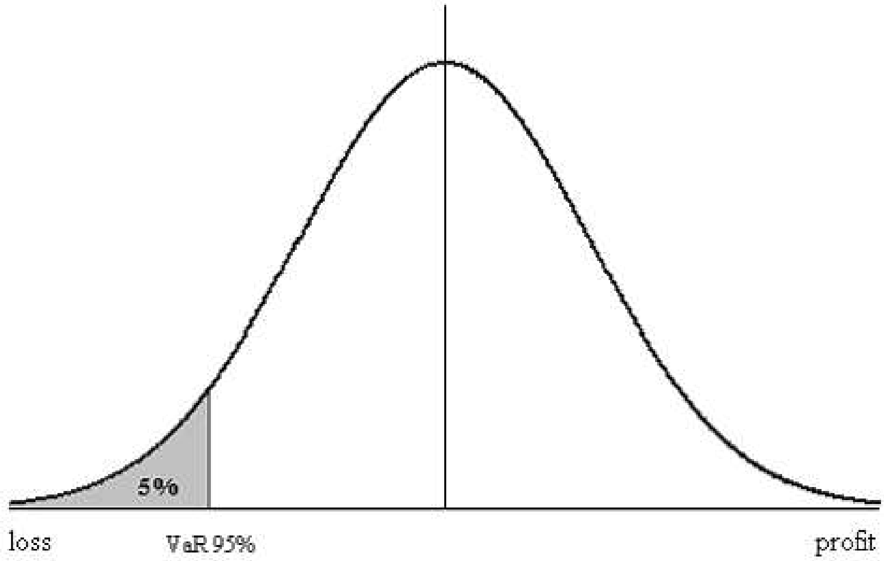

Figure 1: Graphical interpretation of VaR

2.2.2 Criticism on VaR

Not until 1997, with the appearance of *Thinking Coherently* [7] by Artzner et al., it was defined in a clear way what properties a statistic should have in order to be considered a *coherent risk measure.* Artzner et al. (see [8] for a more technical presentation) formulated four axioms that have to be fulfilled by a coherent risk measure. X and Y denote portfolio returns, $\rho(X)$ and $\rho(Y)$ are their risk measures, respectively, and c is an arbitrary constant:

Translation invariance

$$\rho(X + c) = \rho(X) - c \tag{3}$$

Subadditivity

$$\rho(X + Y) \leq \rho(X) + \rho(Y) \tag{4}$$

Positive homogeneity

$$\rho(cX) = c\rho(X) \tag{5}$$

Monotonicity

$$\rho(X) \leq \rho(Y), \text{ if } X \geq Y \tag{6}$$

The translation invariance axiom (3) means that adding cash to the portfolio decreases the risk by the same amount. The axiom of subadditivity (4) ensures that the risk of the total portfolio is not larger than the sum of the risks of its components to reflect the effect of diversification and hedges. Positive homogeneity (5) means that the risk is scaled with the portfolio size. Finally, monotonicity (6) is required to ensure that if the payoff of portfolio X dominates the payoff of portfolio Y, then the risk of portfolio Y cannot be lower than the risk of portfolio X [41]. In simple words, the axioms defining a coherent risk measure means that whenever a portfolio is undoubtedly riskier than another one, it will always have a higher risk value as long as the risk measure is coherent. On the other hand, a measure not fulfilling all axioms might give wrong assessment of relative risks [1]. The most surprising part of the new concept was that VaR, despite its wide acceptance, did not fulfill all axioms of coherence [2]. In fact, VaR fails to meet the characteristic of subadditivity[2], i.e. the risk of a portfolio in terms of VaR may be larger than the sum of risks of its components. The subadditivity condition plays a fundamental role in risk measurement. With non-subadditivity it could be the case that a well diversified portfolio require more regulatory capital than a less diversified portfolio. Thus, managing risk in terms of VaR prevents to add up the VaR of different risk sources and may fail to stimulate diversification (see e.g. [1], [7], [8] or [40]).

The non-subadditivity characteristic of VaR can be demonstrated by a simple example. Suppose that we have two short positions in out-of-the-money binary options. The specific details are shown in Table 1. Each of the options has a 4% probability of a payout of $-\$100$ and a 96% probability of a payout of zero. If we take the VaR at the 95% confidence level, then each of the positions has a VaR of zero. However, if we combine the two positions, the probability of a zero payout falls to less than 95%, and so the VaR of the combined portfolio is less than zero (in this case equal to $-\$100$, see Table 2). The VaR of the combined position is therefore greater than the sum of the VaRs of the individual components, so the VaR is not subadditive.

Table 1: Non-subadditivity: Options positions considered separately

OPTION A		**OPTION B**	
Payout	**Probability**	**Payout**	**Probability**
-\$100	4%	-\$100	4%
0	96%	0	96%
VaR 95% = 0		VaR 95% = 0	

Another criticism on VaR is based on its non-convexity characteristic. The lack of convexity limits its use as a risk measure in optimal portfolio selection for investment purposes. It has been shown [5] that having embedded VaR into an optimization framework, VaR risk managers incur larger losses than

[2] However, VaR is a coherent risk measure when it is based on the standard deviation of normal distributions

Table 2: Non-subadditivity: Options positions combined

COMBINED	
Payout	**Probability**
-\$200	0.16%
-\$100	7.68%
0	92.16%
VaR 95% = -\$100	

non-risk managers in the most adverse states of the world. Moreover, Basak and Shapiro [9] show that an agent facing a VaR constraint may choose a larger exposure to risky assets than in the absence of the constraint. It is also shown in [35] and [36] that the problem of minimizing VaR of a portfolio of derivative contracts can have multiple local minimizers, which will lead to unstable risk ranking. Furthermore, it seems inappropriate to use VaR in practice because of its non-convexity characteristic.

In 1997, when the concept of coherent risk measures first appeared, it became clear that VaR cannot be considered as an adequate risk measure. In spite of this, VaR has been adopted as the main measure of market risk by many financial institutions and has been embraced by risk managers as an important tool in the overall risk management process. The favours of VaR has also been recognised by regulatory authorities. For example, coordinated by the Basel Committee [44], VaR serves for the determination of capital requirements for banks and many national regulatory agencies have adopted the Basel Committee recommendations (e.g. the Swedish Financial Supervisory Authority (Finansinspektionen [45]).

2.2.3 Conditional Value-at-Risk

VaR is often criticized for not taking into account the magnitude of losses when VaR is exceeded. For example, VaR provides no insight into what would happen to a bank if a 1 in 1000 chance event occured. CVaR is often proposed as an alternative to VaR. CVaR is also known as *expected shortfall* [1], *tail VaR* [7] or *mean shortfall* [35]. In the context of continuous distributions (which we assume for simplicity in this paper), for a given confidence level α and holding period t, CVaR is defined as the conditional expectation of the losses exceeding VaR. Hence, in contrast to VaR, CVaR provides additional information of the losses in the tail exceeding VaR.

Mathematically, CVaR is defined by:

$$\mathrm{CVaR}_{\alpha} = \frac{1}{1-\alpha}\int_{-\infty}^{\mathrm{VaR}_{\alpha}} x f_X(x)\mathrm{d}x \tag{7}$$

or equivalently

$$\mathrm{CVaR}_\alpha = E[x | x \leq \mathrm{VaR}_\alpha] \tag{8}$$

where $f(x)$ is the marginal probability function of portfolio returns x over the given time horizon and VaR is calculated over the same time horizon with confidence level α.

A graphical interpretation of CVaR is illustrated in Figure 2. CVaR is the

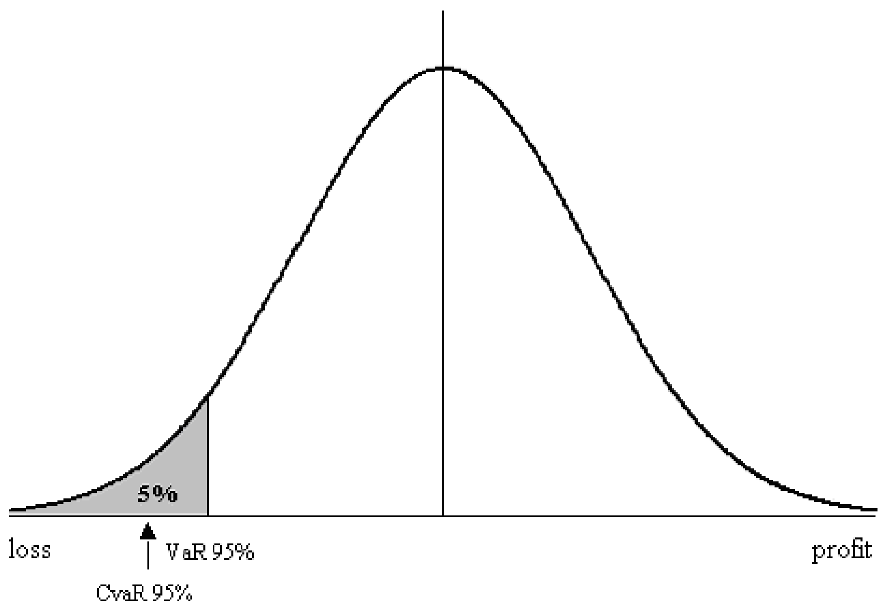

Figure 2: Graphical interpretation of CVaR

expected loss if a tail event does occur, and is therefore graphically located to the left of VaR.

Acerbi and Tasche [3] show that CVaR satisfies the four axioms in section 2.2.2 and, consequently, qualifies as a coherent risk measure. In fact, [4] shows that any coherent risk measure can be represented as a convex combination of CVaRs with different confidence levels. In addition, CVaR is a convex function with respect to portfolio positions, allowing the construction of efficient optimizing algorithms. In particular, it has been shown [42] that CVaR can be minimized using linear programming techniques, which makes many large-scale calculations practical, efficient and stable.[3]

[3] In fact, the superintendent office of financial institutions in Canada has put in regulation for the use of CVaR to determine the capital requirement.

2.3 Estimating VaR and CVaR

There are many ways of estimating VaR (see Duffie and Pan [22] for an overview). Given the return distribution, the calculation of VaR is straightforward and given VaR, the calculation of CVaR is straightforward. Therefore, the challenges of estimating VaR and CVaR are mainly related to the estimation of the return distribution. The approaches can be categorised to parametric and non-parametric methods. Parametric approaches make some assumptions about the return distribution, e.g. the assumption of normality (see section 2.3.1). The distribution assumptions imply model risk, i.e. the risk that there is a discrepancy between the assumed return distribution and the *true* underlying probability distribution [20]. Non-parametric methods base the VaR estimation solely on empirical distributions of returns. A disadvantage is that the estimates are completely dependent on a particular data set. The simplest non-parametric method is called *historical simulation method* (see section 2.3.2).

2.3.1 Delta-Normal Approach

The simplest parametric method is the delta-normal (analytic) approach. Following this approach it is assumed that all asset returns are normally distributed. As the portfolio return is a linear combination of normal variables, it is also normally distributed. The VaR of a portfolio is then calculated using historical (ex ante) means, variances and covariances of the portfolio components. More formally, this can be written as:

$$\mathrm{VaR}_\alpha = \mu - z_\alpha \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij}} = \mu - z_\alpha \sigma_p \tag{9}$$

where w_i and w_j denote the weights of asset i and j in the portfolio of n assets, respectively. σ_{ij} denotes the covariance between returns of asset i and asset j, μ is the mean value of the returns of the portfolio and σ_p is the standard deviation of the total portfolio returns. The parameter z_α is the value of the cumulative normal distribution corresponding to the specific confidence level α, e.g. for the 95% confidence level $z_{95\%} = 1.64$ and for the 99% confidence level $z_{99\%} = 2.33$. Since the holding period is usually short (e.g. ten days) the assumption of a zero mean ($\mu = 0$) is often made. Thus, VaR of a portfolio is simply a multiple of the portfolio standard deviation. After calculating VaR, the calculation of CVaR is straightforward as the expected value of the portfolio losses exceeding VaR.

A major drawback with the delta-normal approach is the exposure to model risk. Even though normal distributions seem to describe the centre of *true* distributions rather well, problems arise when it comes to estimating the tails of distributions. Many empirical studies (see e.g. [17], [25], [26] and [32]) show that the assumption of normally distributed financial returns underestimates VaR. The underestimation becomes more significant when studying securities with heavy-tailed distributions and a high potential for large losses, i.e. that exhibit excess kurtosis [43]. In a similar fashion, Andersen et al. [6] show that accounting for heavy tails makes it possible to increase returns while lowering large risks. These empirical findings are intuitive since heavy tails mean that extreme outcomes are more frequent than what the use of a normal distribution would predict and therefore heavy tails lead to underestimated VaR measures.

In spite of its drawbacks, the delta-normal approach is widely used among risk managers. For example, the RiskMetrics system is based on the parametric delta-normal model.

2.3.2 Historical Simulation

The most common and probably simplest non-parametric method to estimate VaR (and CVaR) is based on historical simulation. The main assumption is that trends of past price changes will continue in the future. The VaR (and CVaR) of a portfolio is then calculated using the percentile of the empirical distribution corresponding to the chosen confidence level. There is no need to estimate distribution parameters such as volatilities and correlation coefficients. The historical simulation method is relatively simple to implement, just keep a historical record of past returns. The method is also free from model risk and makes it possible to accommodate the non-normal distributions with heavy tails that are often found in financial data [25].

The number of past observations to be included in the empirical distribution is often referred to as window size. The choice of window size has a significant impact on VaR measures, especially when using historical simulation [25]. A long window size may include observations that are not relevant to the current situation and may imply a fairly constant VaR measure. A short window size makes the calculations sensitive with respect to abnormal outcomes in the recent past and may imply high variance in VaR measures. The Swedish Financial Supervisory Authority (Finansinspektionen) recommends a window size of at least one year [23].

Many large financial institutions and risk managers compute the VaR of their trading portfolios using the historical simulation approach, e.g. Goldman Sachs [24].

2.3.3 Alternative Approaches

Another widely used approach is the *Monte Carlo simulation*, where a future probability distribution is assumed and the behavior of asset prices is simulated by generating random price paths. The VaR measures can then be determined from the distribution of simulated portfolio values. Monte Carlo frameworks have been shown to provide the best estimates for VaR (see e.g. [31] and [38]). However, at the same time, these models are extremely computer intensive and the additional information that these techniques provide is of most use for the analysis of complex options portfolios.

The *stress testing* method examines the effects of large movements in key financial variables on the portfolio value. The price movements are simulated in line with certain scenarios[4]. Portfolio assets are re-evaluated under each scenario and estimating a probability for each scenario allows to construct a distribution of portfolio returns, from which VaR can be derived.

2.4 Backtesting

Assessing the correctness of VaR models is not an easy task. Since the *true* VaR measures cannot be observed, the evaluation of VaR models must be verified by backtesting. It means that, for a given backtesting period, the estimated VaR measures are compared to the observed returns [12].

2.4.1 Backtesting VaR

There are several possible ways to backtest VaR models (see e.g. [28] and [30]). Typically, the number of times the portfolio loss exceeds VaR is calculated. For each backtesting period the number of violations is calculated. This number of violations divided by the number of observations in the backtesting period gives the violation rate, to be compared to the expected rate of violations. For example, VaR at the 95% confidence level has an expected rate of violations of 5%, and for VaR 99% the expected rate of violations is 1%.

The most widely used test is developed by Kupiec [28]. He examines whether the observed violation rate is statistically equal to the expected one. Under the null hypothesis that the model is adequate, the appropriate likelihood ratio statistic is:

$$L = 2\ln\left(\left(1-\frac{n}{T}\right)^{T-n}\left(\frac{n}{T}\right)^{n}\right) - 2\ln\left((1-q)^{T-n}q^{n}\right) \sim \chi^2_1 \qquad (10)$$

[4] such as movements of the yield curve, changes in exchange rates, etc.

where n is the number of days over a period T that a violation occurred and q is the expected violation rate. Therefore, the risk model is rejected if it generates too many or too few violations.

2.4.2 Backtesting CVaR

To implement a backtesting procedure for CVaR, we need to specify a loss function ρ. A number of different loss functions have been suggested, one of them is proposed by Blanco and Nihle [13]:

$$\rho = \frac{(\text{return} - \text{VaR})^{+}}{\text{VaR}} \tag{11}$$

where $f^{+} = f$ if $f > 0$ and 0 otherwise.

The suggested function gives each tail-loss observation a weight equal to the tail loss divided by the VaR. This ensures that higher tail losses get awarded higher ρ-values. The benchmark for this forecast evaluation procedure is easy to derive. It is equal to the difference between CVaR and VaR, divided by VaR. However, the loss function (11) also has a problem. Because VaR is in the denominator, it is not defined if VaR is zero, and can give mischievous answers if VaR gets close to zero or changes sign.

2.5 Robustness

Most risk measures, such as VaR and CVaR, are defined as functions of the distribution of the considered return. However, since the probability measure describing market events is unknown the distinction between the theoretical risk measure and its estimator allows us to study the relation between the choice of the estimator and the specification of risk measures. In particular, it allows us to consider some natural requirements of the risk measurement procedure. For example, how robust is the result with respect to the data set or with respect to other parameters? Constructing and computing measures of sensitivity allows a quantification of the robustness of VaR and CVaR with respect to the data set and parameters used to compute them. However, VaR and CVaR have completely different properties. Comparing them directly is like comparing apples and oranges. The differences in properties stems from the fact that VaR is an estimate of a percentile in the distribution of returns, i.e. a single point in the distribution. CVaR, on the other hand, is the expected value of returns beyond the VaR percentile, i.e. an estimate that takes all points beyond the VaR percentile into account, though it is condensed into a single scalar estimate. In turn, this means e.g. that CVaR is always less than (or equal to) VaR since CVaR is the expected loss given that the actual return is less than VaR.

For a risk model to be considered robust, it should provide accurate risk forecasts across different assets, time horizons, and confidence levels within the same asset class. Fluctuations in risk forecasts have serious implications for the usefulness of a risk model. However, risk forecast fluctuations have not been well documented. The reason for this is unclear, but the importance of this issue is real. If a VaR value always fluctuates by 30% from one day to the next, it may be hard to sell risk modelling within the firm. Traders are not likely to be happy with routinely changing risk limits, and management does not like to change market risk capital levels too often. Moreover, since VaR is used to regulate market risk capital, a volatile VaR leads to costly fluctuations in capital if the firm keeps its capital at the predicted minimum level. This may severely hinder the adoption of risk models within a firm.

3 Data

In this section we present how we select and collect the data sample used in the thesis. In addition, we give a descriptive overview of the collected data.

For the purpose of this study we use Datastream to gather time series data for equity indices, bond indices, exchange rates and individual stocks. Datastrem is a comprehensive online historical database service provided by Thompson Financial, which is a globally leading supplier of financial information. Data contained in Datastream has been compiled by good faith from sources believed to be reliable. However, Thomson Financial gives no warranty as to its accuracy, completeness or correctness. Balanced against these warnings, however, we believe that Datastream is a practical and reliable source for the type of information used in this study.

Daily prices are collected for the ten year period from May 10, 1996 to May 9, 2006. Ten years of historical data is the longest period available that holds for all variables in the data set. Equity prices are adjusted for dividends, share repurchases and share issues. Non-trading days are excluded from the data set.

The data used in this thesis consists of a variety of major international equity and bond indices as well as major exchange rates and individual stocks. More specifically, we use 42 international equity indices from Europe (ex Sweden), America (ex US), Asia-Pacific and Africa. Most indices were listed on Yahoo Finance [46] as "major world indices". In addition, we use 13 US market equity indices and 39 Swedish market equity indices. Moreover, we use a data sample of 20 major bond indices, 17 major exchange rates and 15 individual international stocks (five stocks from the Stockholm Stock Exchange, five stocks from the MICEX index, which comprises the most liquid Russian stocks and finally five stocks from the Dow Jones Industrial Average index in New York). The analysis is restricted to simplest possible portfolios consisting of a single asset (equity or bond index, exchange rate or individual stock).

An example of a portfolio return distribution over time as well as VaR and CVaR estimates is given for Affärsvärldens Generalindex in Figure (3). As expected CVaR is always less than or equal to VaR.

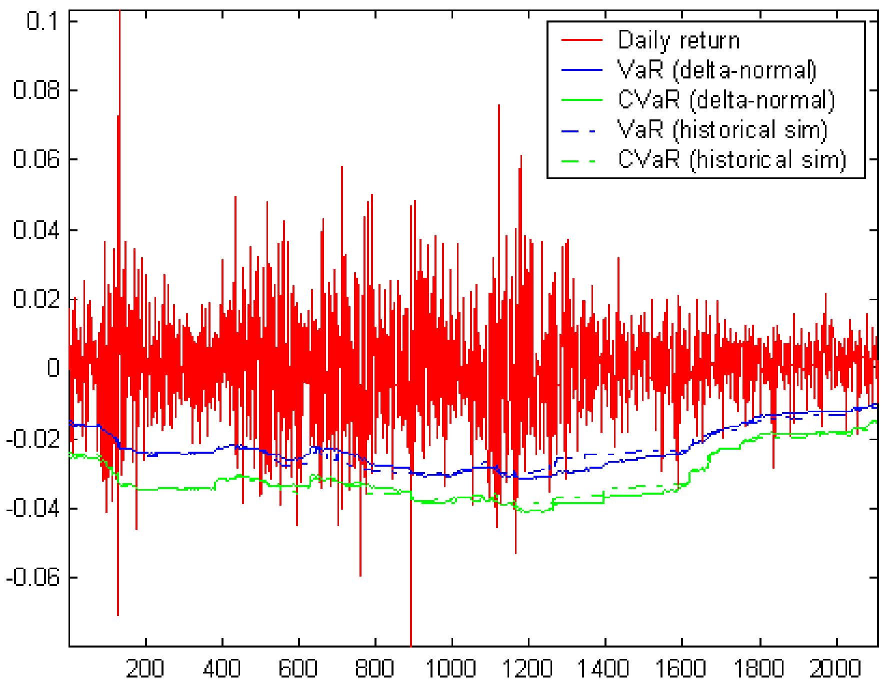

Figure 3: Affärsvärldens Generalindex

4 Hypotheses

In this section, we develop our hypotheses about the robustness of CVaR. We develop two hypotheses, the first about the robustness over all asset classes examined and the second about the difference in robustness between different asset classes.

4.1 Robustness of CVaR

In this thesis we implicitly assume that we compare CVaR to VaR, though a direct comparison is never performed. The reason is that the two measures have completely different properties and as mentioned before we cannot compare apples with oranges.

We hypothesize that CVaR is a more robust risk measure than VaR. The reason is that all values in the tail of the distribution of returns are considered when estimating CVaR, compared to just the number of values for the case of VaR. For example, if the tail consists of the returns -8%, -9.5% and -11%, all these values are taken into account when estimating CVaR. When VaR is estimated, the most important feature about the tail is that it consists of three different (in this example) values. It should be noted, that it is the tail of the distribution that is important for most risk measures, since the tail in some sense defines the risk. The variance of the estimate of a mean (e.g. CVaR) should, intuitively, be less than the variance of the estimation of a single point (e.g. VaR), and therefore the variance should be lower for CVaR compared to VaR. Hence CVaR ought to be a more robust measure of risk.

Furthermore, a formal test is performed where we test whether CVaR is an unbiased measure of the conditional return, giving that the return is less than VaR. We perform the test by testing the null hypothesis H_0 that the empirical measures, respectively, are zero on average against the alternative hypothesis H_1 that they differ from zero. The empirical measures are by design equal to zero if CVaR is an unbiased estimate of the conditional return.

4.2 Asset Class Difference

The estimation of CVaR, as well as of VaR, is always based on some kind of model which makes use of ex ante data (cf. section 2.3). Inherent in the model is some kind of assumptions about the characteristics of the underlying data and relevant model parameters are estimated based on the ex ante data.

As a result, the VaR and CVaR estimates get better the better the model assumptions agree with the data at hand. It is perhaps more intuitive to think about it the other way around. When you construct a model, you try and design the appropriate model assumptions based on your believes of the real data. Hence the data samples that are most similar to the model assumptions will render the most stable CVaR estimates. In our case, we estimate CVaR based on two different methods, the first assuming normally distributed returns (the delta-normal approach, see 2.3.1) and the second assuming that the distribution of returns in the ex ante period is representative for the distribution of future returns (the historical simulation method, see 2.3.2). This leads to that the asset class that, on average, has returns that are most similar to a normal distribution will seem to be the most robust asset class, in terms of CVaR robustness, when CVaR is estimated using the delta-normal method. On the other hand the asset class that, on average, is most constant over time, will seem to be the most robust asset class when CVaR is estimated using the historical simulation method. It is without further studies impossible to say what asset class that has returns that are most similar to a normal distribution or what asset class that has a pattern of returns that is the most constant over time.

We will test the null hypothesis H_0 that the different asset classes, pair-wise, on average have the same value of the empirical measure CRV (cf. section 5.3) against the alternative hypothesis H_1 that they differ. If CVaR is a robust measure of risk, there should be no difference over the different asset classes in terms of robustness (though the actual risk will of course differ) if CVaR is an unbiased risk measure for those particular asset classes.

5 Methodology

In this section we describe the methodology. First, the procedure how the study is performed is described. Thereafter, the empirical measures used in the study are described. Finally, the empirical tests performed are described.

5.1 Returns

Throughout the thesis we calculate the daily returns as:

$$r_t = \ln \frac{P_t}{P_{t-1}}$$

where r_t is the daily return, P_t is the closing price on day t and P_{t-1} is the closing price on day $t-1$. In other words, we follow the standard in financial analysis and use log-returns.

5.2 Procedure

By using data in the ex ante time period, we calculate VaR and CVaR. VaR is calculated using one of several possible methods, which are described in section 2.3. CVaR is calculated accordingly.

Thereafter, the accumulated return τ_H days after the end of the ex ante period is observed, where τ_H is the so called VaR horizon which is the hypothesized holding period. If the return is less than VaR, the event is counted. This is later on used to evaluate VaR. On average, we should observe $100(1-\alpha)$ returns less than VaR if VaR is a good measure and α is the confidence level of VaR. Furthermore, the empirical measures to evaluate CVaR are calculated. The calculation of these empirical measures are described in section 5.3.

The next step is to move the ex ante window one day forward and repeat the steps described above, starting with calculating updated values of VaR and CVaR for the new ex ante time period. This procedure is than repeated until the end of the data file.

This algorithm is then repeated for all assets, methods of calculating VaR and different parameters under study. The parameters that can be varied are the VaR horizon τ_H, the ex ante window length and the confidence level. The resulting empirical measures can also be averages over different parameters.

In the next subsection, the different empirical measures for evaluating the robustness of CVaR are described.

5.3 Empirical Measures

In order to derive the answers to our questions, we develop three different empirical measures that we believe will capture the behavior of CVaR as well as possible. The empirical measures are CVaR relative to VaR (CRV), adjusted CVaR relative to VaR (adjCRV) and CVaR relative to return (CRR). The measures are described in the subsequent subsections.

The reason that we develop our own empirical measures is that there is no appropriate standard measure in the literature, at least to our knowledge. The reason that we develop more than one measure is that there is no single good measure that captures all behaviors of the phenomena under study. Hence, we develop CRV and CRR as complements to each other. The adjCRV consists of a slight modification of CRV which makes it directly comparable to CRR. One necessary condition that all empirical measures should fulfill in order to be good measures is that they should point in the same direction every time.

5.3.1 CVaR Relative to VaR (CRV)

For each sample of returns[5], the ex ante VaR and CVaR are calculated based on data in the ex ante period. By definition, CVaR is always less than VaR. The value $\rho_1 = \frac{\text{CVaR} - \text{VaR}}{\text{VaR}}$, which measures how much smaller CVaR is compared to VaR, is recorded. As a second step, the actual return is examined. If the return is larger than VaR, nothing is recorded and we continue with the next step in the algorithm and form a new ex ante period. However, if the return is abnormal and negative, i.e. less than VaR[6], the value $\rho_2 = \frac{\text{return} - \text{VaR}}{\text{VaR}}$, which measures how much smaller the return is compared to VaR, is recorded. As a last step the difference $\rho_3 = \rho_2 - \rho_1$ is calculated. This is a measure of the difference between CVaR and the abnormal negative return. The unit is the somewhat non-intuitive "percent of VaR". Thereafter, we form a new ex ante period one step forward in time and the algorithm is repeated from start until all samples of returns have been examined. An example can be used to illustrate CRV. If the ex ante VaR and CVaR are -10% and -15%, respectively, and the return is -14%:

$$\begin{aligned} \rho_1 &= \tfrac{\text{CVaR}-\text{VaR}}{\text{VaR}} = \tfrac{(-15)-(-10)}{(-10)} = 50\% \\ \rho_2 &= \tfrac{\text{return}-\text{VaR}}{\text{VaR}} = \tfrac{(-14)-(-10)}{(-10)} = 40\% \\ \rho_3 &= \rho_2 - \rho_1 = 40\% - 50\% = -10\% \end{aligned}$$

[5] e.g. each trading day if the VaR horizon is one trading day

[6] inherent in the VaR and CVaR concepts are that returns are considered abnormal (and negative) when they are less than VaR for the relevant confidence level

The interpretation is that on this occasion, we see an abnormal negative return (since the return is less than VaR) and the return is 10% of VaR smaller than CVaR (taking the sign into account). This means that the return is larger than CVaR, since VaR has a negative sign.

This leaves us with a number of different ρ_3 for each asset. There should be approximately $1 - \alpha$ times the number of samples number of ρ_3. All ρ_3 related to a specific asset are aggregated into a scalar measure, the mean.

An advantage with CRV compared to CRR (described below) is that CRV by design is adjusted for different volatilities of the different assets. A similar measure was also suggested by Dowd in [21]. He notes that this is measure is problematic since it is not defined in the case when VaR is zero and can be mischievous if VaR is close to zero.

5.3.2 Adjusted CVaR Relative to VaR (adjCRV)

adjCRV is similar to CRV. The only difference is that ρ_3 is multiplied by the average VaR which results in a unit of adjCRV which is more easily interpreted. The unit simply becomes "percent" (of the original asset value). Another way to look at it is to view adjCRV as a linearly scaled version of CRV which makes it directly comparable with CRR. Notice that due to our definition of VaR, CRV and adjCRV will on most occasions have different signs. This is due to the fact that VaR is negative on average.

To illustrate, we continue with our example above. We got so far that we identified an abnormal return of 10% of VaR smaller than CVaR. But this is just simply 10% of -10% which equals -1% (assuming an average VaR of -10%). Hence the return is -1% smaller than CVaR, i.e. 1% larger than CVaR (-14% compared to -15%).

Again, we are left with a number of different adjusted ρ_3 for each asset, and again the ρ_3 are aggregated into the scalar measure the mean.

An advantage with adjCRV is that it makes the empirical measures CRV and CRR directly compareable.

5.3.3 CVaR Relative to Return (CRR)

The last measure, CRR, compares CVaR to the return in the cases where the return is less than VaR. For each sample of returns, the ex ante VaR and CVaR are calculated just as before. The return is then compared to VaR. If the return is larger than VaR, nothing is recorded and we continue with the next step in the algorithm and form a new ex ante period. However, if the return is abnormal and negative, i.e. less than VaR, the difference return minus CVaR is recorded. This is a measure of the difference between the abnormal negative return and CVaR. The unit is "percent" (of the original asset value). Hence, the numbers of CRR are directly comparable to the numbers of adjCRV. This was the rationale behind the adjustment of CRV in the first place. Thereafter, we form a new ex ante period one step forward in time and the algorithm is repeated from start until all samples of returns have been examined.

Using the same figures as in the previous example, CRR becomes $-14\% - -15\% = 1\%$. Hence, the return is 1% larger than CVaR, which agrees with the result for adj CRV. Though adjCRV and CRR reached exactly the same value this time, it should be noted that this is not the case in general.

Also in the case of CRR, the different values for each sample of returns for each asset are aggregated into a scalar measures, the mean.

An advantage with CRR compared to CRV is that its unit is easily interpreted.

5.4 Empirical Tests

Two different tests are performed to evaluate CVaR. We will call them the *CVaR Robustness* test and the *Asset Class Difference* test. They are described in the subsequent subsections.

5.4.1 CVaR Robustness

The values of the different empirical measures described in section 5.3 will be analyzed based upon their respective magnitudes. A simple formal statistical test will be performed where we check whether the means of the empirical measures are zero or not. If the measures are zero, this indicates that CVaR is a good risk measure on average.

On the other, also the variation of the empirical measures must be taken into account when the robustness of CVaR is evaluated. A good measure of the robustness is the Inter-Quantile Range (IQR), which measures the difference between two quantiles in the distribution of an empirical measure. $\text{IQR}_{0.95-0.05}$ measures e.g. the difference between the five and ninety five percentile. This measure gives an indication of the spread of the empirical measure, and hence also an indication of the robustness of CVaR. If CVaR is a robust risk measure, IQR should be small. However, no formal statistical tests based upon the variation or IQR will be performed. The reason is that a straight-forward test that would give insight to the problem is not easily constructed. As a complement in helping us characterizing the loss, we will also study the conditional return distribution given that the return is less than VaR and relate this to the estimated VaR and CVaR.

5.4.2 Asset Class Difference

In this test, the different asset classes are ranked according to the empirical measure CRV. The reason that only CRV is considered in this test is that it is the only empirical measure that is adjusted for the volatility of the underlying asset (cf. section 5.3). By ranking the different assets, we hope to be able to draw the conclusion whether CVaR is a robust risk measure by examining the difference in CRV between the different asset classes. If CVaR is a robust measure of risk, it should be transparent to the underlying type of asset, in the sense that a measure of robustness should not be different for different asset classes. Of course, the value of CVaR itself will vary significantly of different assest since e.g. T-bills are less risky than a stock in a mineral company listed on the Moscow Stock Exchange. A simple 2-sample t-test where the variances of the two populations are not assumed equal is performed to investigate the pair-wise difference in CRV between the different asset classes.

6 Empirical Findings

In this section the results of the study are presented. In the first part of the study, all asset classes are treated jointly, whereas in the second part, they are treated individually. The different asset classes considered in this thesis are stock indices, bond indices, exchange rates and individual stocks.

It should be noted that our primary goal is to determine whether CVaR is a robust risk measure or not and to some extent characterize the conditional return, given that the return is less than VaR. It is not our intention to give exact numerical answers to these questions.

6.1 Parameters

As described in section 2.3, VaR and CVaR can be estimated using different methods. In all methods, some parameters always have to be decided beforehand. In our case, the parameters are the ex ante estimation window length, the confidence level of VaR and CVaR and the horizon.

The *ex ante estimation window length* is chosen to 250 or 500 trading days, corresponding to approximately one year and two years, respectively. The ex ante estimation window is used to estimate the model parameters, e.g. the standard deviation of the returns if that is an input to the model. Lambadiaris et al. [29] uses an ex ante length of 100 or 252 trading days. They conclude that a longer estimation window is usually better. The reason that they do not use a longer ex ante window is a restriction in the number of samples. The Swedish Financial Supervisory Authority (Finansinspektionen) [23] suggests an ex ante window length of more than one year.

The *confidence level of VaR and CVaR* is chosen to 95% or 99%. These are the most common values in the literature. Furtehrmore, the Swedish Financial Supervisory Authority (Finansinspektionen) [23] suggests a confidence interval of at least 99%.

The VaR and CVaR *horizon* is chosen to be one day. This is a common value in the literature. The horizon is the same as the hypothetical holding period and hence the relevant return is the return during the VaR horizon period. In our case, where we study market risk, the VaR horizon is typically one trading day. However, the Swedish Financial Supervisory Authority (Finansinspektionen) [23] suggests a VaR horizon of ten days. On the other hand, they also say that it is equally good to perform all calculations assuming a one-day horizon and then as a final step calculate the final ten-day horizon VaR (or CVaR) value from the one-day horizon value through a simple transformation.

6.2 CVaR Robustness

6.2.1 Possible Bias

Part of the empirical results are presented in Tables 3 - 6. We choose to present the measure CRR only, since all empirical measures point in the same direction. The complete set of empirical results is found in Appendix A. Since CVaR is closely related to VaR, we also do a back-testing of VaR (see Table 8).

The entries in Tables 3-4 should be interpreted as follows: Consider e.g. the second line. The first column tells us that the parameters in this case are an ex ante window length of 500 trading days and a confidence level of 95%. The second column in Table 3 tells us that the average CRR over all assets is -0.009%. This means that if CVaR is e.g. -8%, the conditional return, given that the return is less than VaR, is -8.009% on average. Columns 3-6 in Table 3 and columns 2-3 in Table 4 are different measures of the deviation of CRR about its mean, where Q stands for quantile and IQR for Inter-Quantile Range. The last two columns in Table 4 show the result of the test whether the mean CRR is equal to zero, and the associated t-statistic is given in column 4 in Table 4.

Table 3: Empirical results of CRR using the delta-normal method

ex ante length / confidence level	mean CRR	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.95}$	$Q_{0.975}$
500 / 95%	-0.009	-0.25	-0.20	0.14	0.17
500 / 99%	-0.029	-0.90	-0.44	0.31	0.37
250 / 95%	-0.07	-0.30	-0.26	0.067	0.093
250 / 99%	-0.13	-1.15	-0.71	0.16	0.17

Table 4: (cont.) Empirical results of CRR using the delta-normal method

ex ante length / confidence level	$IQR_{0.95\text{-}0.05}$	$IQR_{0.975\text{-}0.025}$	t-stat	$H_0^{95\%}$	$H_0^{90\%}$
500 / 95%	0.34	0.42	-0.083	acc	acc
500 / 99%	0.74	1.27	-0.068	acc	acc
250 / 95%	0.32	0.39	-0.43	acc	acc
250 / 99%	0.87	1.32	-0.297	acc	acc

We note that the hypothesis H_0 that the empirical measures are equal to zero, and hence that CVaR is an unbiased estimate of the conditional return, given that the return is less than VaR, is accepted in most cases. It is rejected only in two cases (see Appendix A) at a rather low 90% confidence level.

Table 5: Empirical results of CRR using the historical simulation method

ex ante length / confidence level	mean CRR	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.95}$	$Q_{0.975}$
500 / 95%	-0.035	-0.30	-0.23	0.11	0.14
500 / 99%	-0.29	-1.81	-1.52	0.17	0.24
250 / 95%	-0.11	-0.42	-0.30	0.002	0.009
250 / 99%	-0.43	-1.78	-1.30	-0.031	-0.004

Table 6: (cont.) Empirical results of CRR using the historical simulation method

ex ante length confidence level	$IQR_{0.95\text{-}0.05}$	$IQR_{0.975\text{-}0.025}$	t-stat	$\mathrm{H}_0^{95\%}$	$\mathrm{H}_0^{90\%}$
500 / 95%	0.34	0.45	-0.295	acc	acc
500 / 99%	1.69	2.05	-0.419	acc	acc
250 / 95%	0.30	0.43	-0.918	acc	acc
250 / 99%	1.27	1.77	-0.725	acc	acc

In both cases, the ex ante window length is the shorter of the possibilities, which agrees with the results of Lambadiaris et al. [29], who concluded that a longer ex ante length gives a better estimate. It can also be noted that both cases happen for the historical simulation method.

The findings support that CVaR is an unbiased estimate. On the other hand, the mean of CRV is always positive and the means of adjCRV and CRR are always negative, which could be interpreted as if CVaR is a biased estimate. However, this is most probably due to the fact that the same methods and the same data are used over and over again when CVaR is estimated, only the parameters of the estimator change.

To further investigate whether CVaR is an unbiased measure, we study CRR of the individual assets and perform a similar test as above to test whether CRR is equal zero or not. The null hypothesis is $H_0 : CRR = 0$. The test is two-sided and performed on a 95% confidence level. The summary of the results are shown in Table 7. As can be seen, CVaR seems to be an unbiased estimate. The parameters of the CVaR estimator are in this case an ex ante window length of 500 trading days and a confidence level of 95%. The rationale behind this choice is that these parameters seem to render the most stable CVaR estimates, as shall be shown subsequently.

Next, in Table 8 we perform a backtesting of VaR. In the table the proportions of returns less than VaR are stated as well as the corresponding p-values calculated from the binomial-test described in section 2.4.1. As we can see, the null hypothesis of binomially distributed VaR values can be

Table 7: Summary of the tests where we study whether CRR for individual assets are equal to zero. The table shows the number of assets in each asset class for which H_0 is rejected (percent of times within parenthesis)

Asset class	# assets	delta-normal	historical simulation
Stock indices	100	1 (1%)	4 (4%)
Stocks	14	0	0
Bonds	19	0	1 (5.3%)
Exchange rates	0	0	0

rejected at a 5% significance level in three of the cases. All rejections refer to the delta-normal approach. It seems like the historical simulation does a better job in estimating VaR values, in fact, the backtesting shows really encouraging results when using the historical simulation method. A possible explanation could be that the assets in the data sample show fat-tail properties and, hence, do not exhibit the normality characteristic assumed in the delta-normal approach.

Table 8: Back-testing of VaR. Violation ratios and p-values

ex ante length / confidence level	500 / 95	500 / 99	250 / 95	250 / 99
Delta-normal method	4.1 **(0.0283)**	1.5 **(0.0233)**	4.5 (0.1899)	1.6 **(0.0043)**
Historical simulation	4.7 (0.4595)	0.9 (0.6760)	5.2 (0.6204)	1.3 (0.2141)

6.2.2 Robustness

Furthermore, the IQR indicates that CVaR is a robust measure of risk. It can also be noted that the IQR increases as the confidence level increases. This is most probably due to the fact that there are fewer samples in this case to base the estimate on, compared to the lower confidence level. On average, there are five times as many samples of the return being less than VaR at the 95% confidence level compared to the 99% confidence level.

Considering only $\text{IQR}_{0.95-0.05}$ for now, the largest value is found for the historical simulation method with an ex ante length of 500 days and a confidence level of 99% (see Appendix A). Here, we only take adjCRV and CRR into account since theses are the only empirical measures where the unit is *percent of the original asset value*. The interpretation of this value would be as follows: The 5-percentile is -1.66% and the 95-percentile is 0.15%. If CVaR is, say, -8%, the true conditional return, given that the return is less than VaR would be between -9.66% and -7.85%, if "true" is interpreted as lying between the 5 and 95 percentiles. This might not seem as a tight interval, but it must be considered that this is the worst outcome out of all the cases. Using a confidence level of 95% instead, the $\mathrm{IQR}_{0.975-0.025}$ is always less than 0.47%, no matter what estimation method that has been used and the ex ante window length. Moreover, in this case we consider a tougher IQR.

A numerical example might be illustrative. Consider the case of an ex ante length of 500 days, a confidence level of 95% and the adjCRV measure. The CVaR estimation method is the delta-normal-method. In this case $\mathrm{IQR}_{0.975-0.025} = 0.45\%$ and the 5-percentile is -0.25% and the 95-percentile is 0.10%. This means that if CVaR is, say, -8%, the true conditional return, given that the return is less than VaR would be between -8.25% and -7.90%, if *true* is interpreted as lying between the 2.5 and 97.5 percentiles.

The best way to get an opinion whether CVaR is a robust measure is perhaps to study the entire distribution of the empirical measures in the form of the cumulative distribution function (CDF). They are presented in Appendix B.

6.2.3 Conditional Return Distributions

As a complement in helping us characterizing the conditional return, i.e. given that the return is less than VaR, the conditional return normalized with respect to VaR is plotted in Figure 4. In the figure, VaR is estimated with the delta-normal method, but the results for the historical simulation method are similar. The value "1" on the abscissa represents normalized Var, and hence a value of "2" means that the return on this occasion was 200% of VaR.

In appendix C (Figures 13 - 22) some additional conditional distribution, normalized with respect to $\kappa(t, \alpha) = \mathrm{VaR}(t, \alpha) - \mathrm{CVaR}(t, \alpha)$ are presented. This normalization makes returns over all assets and over time directly comparable. This makes returns over time and over all assets directly comparable.

6.3 Asset Class Difference

Another way to test the robustness of CVaR is to compare the empirical measure CRV over different asset classes. If CVaR is robust, CRV should not differ over different asset classes, if CVaR is an unbiased risk measure for those particular asset classes. The test is performed as a difference-in-mean test where the null hypothesis H_0 that there is no difference between different asset classes is tested against H_1 that there is a difference in the mean of CRV between different asset classes. The number of degrees of

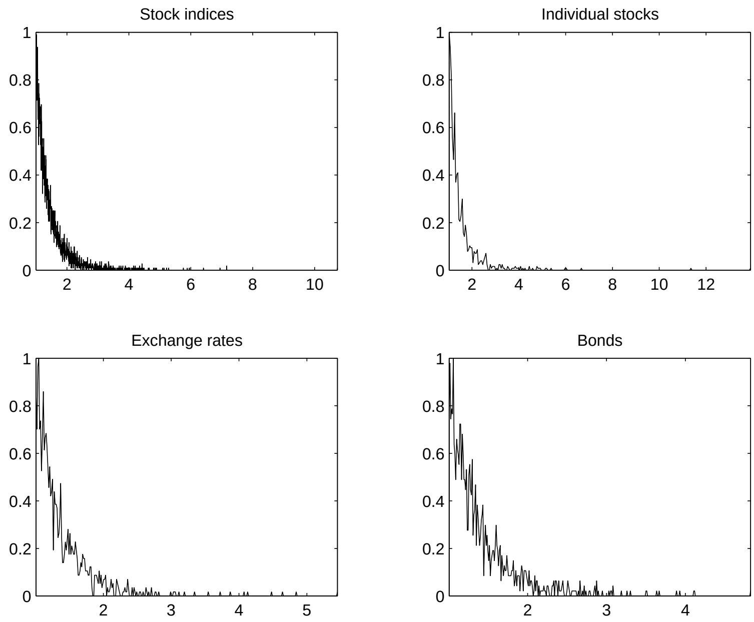

Figure 4: Conditional returns relative to VaR. Ex ante length = 500 days, confidence level = 95%.

freedom (DF) is approximated by the number of assets in the asset class with the lowest number of assets, minus one.

The result of the test is presented in Tables 9 - 10. In the tables, the t-statistic and DF are presented. The critical values are 2.1009 and 2.1604 at the 95% confidence level for 18 and 13 df, respectively, and 1.7341 and 1.7709 at the 90% confidence level for 18 and 13 df, respectively. In the tables, a t-statistic above the critical value at the 95% confidence level is indicated with boldface typesetting.

Table 9: Difference-in-mean test. Delta-normal method. The ex ante length in days and the confidence level in percent are given in the table. ex-rates = exchange rates

Assets	t-stat 500 / 95	t-stat 500 / 99	t-stat 250 / 95	t-stat 250 / 99	DF
stock indices - bonds	-1.483	-1.205	-0.016	-0.534	18
stock indices - ex-rates	-0.191	-0.787	**5.455**	**2.317**	13
bonds - ex-rates	1.426	0.614	1.233	1.546	13
stocks - ex-rates	1.764	1.763	**3.125**	**2.488**	13
stocks - stock indices	1.883	2.126	0.517	1.559	13
stocks - bonds	-0.414	1.283	0.202	0.777	13

Table 10: Difference-in-mean test. Historical simulation method. The ex ante length in days and the confidence level in percent are given in the table. ex-rates = exchange rates

Assets	t-stat 500 / 95	t-stat 500 / 99	t-stat 250 / 95	t-stat 250 / 99	DF
stock indices - bonds	-1.939	-0.706	**3.790**	-1.187	18
stock indices - ex-rates	-0.671	-0.041	**4.214**	1.375	13
bonds - ex-rates	0.687	0.632	-0.696	2.006	13
stocks - ex-rates	0.894	1.843	**3.092**	2.450	13
stocks - stock indices	1.569	2.011	0.375	1.885	13
stocks - bonds	-0.250	1.187	**3.110**	1.039	13

The conclusion that can be drawn from the results so far is that there seems to be a significant difference in terms of the CRV between different asset classes in the case of an ex ante window length of 250 days. On the other hand, when the ex ante window length increases to 500 days, most of the differences seem to disappear. This can be interpreted as the robustness of CVaR increases when the ex ante window length increases, a conclusion that we have touched upon before. Hence, the CVaR robustness seems sensitive to a short ex ante window.

Furthermore, the significant difference between the asset classes might be interpreted as there is a bias in the CVaR estimate. Whether this is true or not is hard to tell. If we analyse what lies beneath the numbers, we find that one part of the explanation to the difference are few individual stocks with extreme movements, mainly from the Moscow stock exchange. These stocks have a significant impact on the average CRV for the stock category, since the number of individual stocks in this study is rather limited. Hence, another interpretation of the results is that CVaR is not a good risk measure for certain asset classes, in this case individual stocks, due to the return distribution of that asset class. This is also in line with the findings in section 6.2.1 where we saw that CVaR sometimes seem to be a biased estimate for individual stocks. However, if there is a bias with a 250 day ex ante window length, it seems to go away as the ex ante window length increases to 500 days.

7 Concluding Remarks

In this section, we summarize our findings and provide some suggestions for further research.

To start with, it seems plausible that CVaR is an unbiased estimate of the conditional return, given that the return is less than VaR, with the possible exclusion of individual stocks.

Secondly, our results indicate that it is possible to construct a stable CVaR. The results suggest that the key is to choose the CVaR estimating parameters carefully. This means choosing a confidence level that is not too high, which we believe is due to the fact that a certain number of samples is needed in the ex ante window to estimate the model parameters accurately. It also means that the ex ante window has to be chosen long enough, probably due to the same reason.

Hence, as the confidence level of VaR increases, the robustness seems to decrease. This is a problem, since a high confidence level of VaR is usually wanted. A 95% confidence level would mean that we consider approximately one trading day per month as being abnormal. A 99% confidence level would mean that we consider approximately two to three trading days per year as being abnormal. When the confidence level increases, the number of relevant samples in the ex ante window decreases so that the CVaR estimate gets worse. This can probably to some extent be compensated by extending the ex ante window length, but this might not be possible due to practical reasons, since there often is a lack of relevant historical data. Even if there were enough historical data available, it might not be representative due to its age.

To end this section, it should be stated that a risk measure is a statistical measure, and hence we expect it to be correct only on average.

In the next section, some further research subjects are suggested.

7.1 Suggestions for Further Research

The results indicate that the robustness of CVaR increases as the ex ante window length increases and the confidence level decreases. This might be due to the fact that the number of historical samples in the ex ante window increases, which renders a better CVaR estimate. It would be interesting to find out approximately how many samples are needed for a good CVaR estimate and if a longer ex ante window length can be directly traded for a higher confidence level in terms of robustness. Or is old historical data less useful than more recent data? What is the trade-off between recent and older data?

Inherent in the CVaR estimation process is the estimation of VaR. This might transfer some of the robustness issues of VaR onto CVaR. It would be interesting to try and isolate the evaluation of the robustness of CVaR from VaR in some way. In our study, we evaluate CVaR every time the return is less than VaR, but since there are issues with VaR, this might not be the best thing to do. One possible alternative might be to evaluate CVaR for e.g. the worst five percent of the returns, if the confidence level is 95 %.

References

[1] Acerbi, C; Nordio, C; Sirtori, C. Expected shortfall as a tool for financial risk management. *Technical report*, 2001.

[2] Acerbi, C; Tasche, D. Expected shortfall: A natural coherent alternative to Value at Risk, *Working paper*, 2001.

[3] Acerbi, C; Tasche, D. On the coherence of expected shortfall, *Working Paper*, Department of Mathematics, TU-Munchen, 2001.

[4] Acerbi, C. Spectral measures of risk: A coherent representation of subjective risk aversion, *Working Paper*, 2002.

[5] Alexander, G.J; Baptista A.M. A Comparison of VaR and CVaR Constraints on Portfolio Selection with the Mean-Variance Model. *Management Science*, Vol 50, No 9, September 2004, pp. 1261-1273.

[6] Andersen, J; Sornette, D. Have your cake and eat it too: increasing return, *Working Paper*, 1999.

[7] Artzner, P; Delbaen, F; Eber, J.M; Heath, D. Thinking coherently, *Risk*, 10:68-71, 1997.

[8] Artzner, P; Delbaen, F; Eber, J.M; Heath, D. Coherent measures of risk, *Math Finance*, 9(3), pp. 203-228, 1999.

[9] Basak, S; Shapiro, A; Value-at-Risk-Based Risk Management: Optimal Policies and Asset Prices. *Review of Financial Studies*, Vol 14, pp. 371-405, 2001.

[10] Basel Committee on Banking Supervision, Amendment to the capital accord to incorporate market risks, *Basel Committee Publications*, Bank for International Settlements, January 1996, p. 1.

[11] Basel Committee on Banking Supervision, International Convergence of Capital Measurement and Capital Standards, *Basel Committee Publications*, Bank for International Settlements, November 2005, p. 147.

[12] Best, P; *Implementing Value At Risk*, John Wiley & Sons, 1999.

[13] Blanco, C; Ihle, G. How good is your VaR? Using backtesting to assess system performance, *Financial Engineering News*, August 1998.

[14] Bodie; Kane; Marcus. *Investments*, 6th edition, McGraw-Hill, New York, 2005.

[15] Boyle, P; Hardy, M; Vorst, T; Life after VaR, *The Journal of Derivatives*, Fall 2005, pp. 48-55.

[16] Citibank, Annual Report, 2005, p.38.

[17] Danielsson, J; de Vries, C. Value-ar-Risk and Extreme Returns, *FMG Discussion Paper*, No 273, Financial Markets Group, London School of Economics, 1997.

[18] Daníelsson, J; Keating, C; Shin, H.S; Muennich, F; Goodhart, C. An Academic Response to Basel II. *FMG Special Papers*, sp130, Financial Markets Group, 2001.

[19] Daníelsson, J; de Vries, C.G; Jorgensen, B.N; Mandira, S. Comparing Downside Risk Measures for Heavy Tailed Distributions, *FMG Discussion Papers*, dp551, Financial Markets Group, 2005.

[20] Dowd, K; Beyond Value At Risk: The New Science of Risk Management, John Wiley & Sons, 1998.

[21] Dowd, K. Retrospective assessment of Value-at-Risk, John Wiley & Sons, 1999.

[22] Duffie, D; Pan, J. An Overview of Value at Risk, *Journal of Derivatives*, Vol 4, pp. 7-49, 1997.

[23] *Intern VaR-modell för beräkning av kapitalkrav för marknadsrisk*, Finansinspektionen, 2004-04-26.

[24] Goldman Sachs, Annual Report, 2005, p.51.

[25] Goorbergh, R; Vlaar, P. Value-at-Risk Analysis of Stock Returns Historical Simulation, Variance Techniques or Tail Index Estimation?, *De Nederlandsche Bank Staff Reports*, March 1999, Nr. 40.

[26] Hendricks, D. Evaluation of value-at-risk models using historical data, *Economic Policy Review*, Federal Reserve Bank of New York, 2 April 1996.

[27] JPMorgan, Annual Report, 2005, p.76.

[28] Kupiec, P.H. Techniques for Verifying the Accuracy of Risk Measurement Models, *Journal of Derivatives*, Vol 3, pp. 73-84, 1995.

[29] Lambadiaris, G; Papadopoulou, L; Skiadopoulos, G; Zoulis, Y. VaR: History or simulation. *Risk Magazine* September 2003, pp. 123-126.

[30] Lopez, J.A. Methods for Evaluating Value-at-Risk Estimates. Federal Reserve Bank of New York, *Economic Policy Review*, Vol 2, pp. 3-17, 1999.

[31] Lucas, A; Klaassen, P. Extreme returns, downside risk, and optimal asset allocation. *Journal of Portfolio Management*, Vol 25, pp. 71-79, 1998.

[32] Mahoney, J. *Empirical-based versus Model-based Approaches to Value-at-Risk: An Examination of Foreign Exchange and Global Equity Portfolios*, Proceedings of a Joint Central Bank Research Conference, Board of Governors of the Federal Reserve System, pp. 199-217, 1996.

[33] Markowitz, H.M. Portfolio Selection, *Journal of Finance*, Vol 7, No 1, pp. 77-91, 1952.

[34] Martín Mato, M.A. Classic and modern measures of risk in fixed-income portfolio optimization. *The Journal of Risk Finance*, Vol 6, No 5, pp. 416-423, 2005.

[35] Mausser, H; Rosen, D. Beyond VaR: From measuring risk to managing risk, *ALGO Research Quarterly*, 1:5-20, 1999.

[36] McKay, R; Keefer, T.E. *VaR is a dangerous technique.* Corporate Finance Searching for Systems Integration Supplement, September:30, 1996.

[37] Olszewski, Y. Building a Better Fund of Hedge Funds: A Fractal and α - Stable Distribution Approach, *Working Paper*, draft #4, August 2005.

[38] Pritsker, M. Evaluating value at risk methodoligies: Accuracy versus computational time, *Journal of Financial Services Research*, Vol 12, pp. 201-242, 1997.

[39] J.P.Morgan/Reuters, *RiskMetrics - Technical Document*, New York, 1994.

[40] Rootzén, H; Klüppelberg, C. A single number can't hedge against economic catastrophes. *Ambio*, 28(6), pp. 550-555, Royal Swedish Academy of Sciences, 1999.

[41] Smithson, C. Beyond VAR, *Risk*, December 2000, Vol 13, No 12, pp. 85-87.

[42] Uryasev, S; Rockafellar, R.T. Optimization of Conditional Value-at-Risk, *The Journal of Risk*, Vol 2, No 3, 2000, pp. 21-41.

[43] Yamai, Y; Yoshiba, T. Comparative Analysis of Expected Shortfall and Value-at-Risk: Their Validity under Market Stress, *Working Paper*, October 2002.

[44] Basel II: Revised international capital framework, Bank for International Settlements
http://www.bis.org/publ/bcbsca.htm

[45] Kapitaltäckning (Basel II), Swedish Financial Supervisory Authority (Finansinspektionen)
http://www.fi.se/Templates/ListPage____2914.aspx

[46] Yahoo Finance, Major world indices
http://uk.finance.yahoo.com/m2

A Empirical Findings - CVaR Robustness

In this appendix, we present the numerical results discussed in section 6.

Table 11: Empirical results using the delta-normal method. Ex ante length = 500 days, confidence level = 95%

	mean	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.95}$	$Q_{0.975}$
CRV	1.32	-5.52	-4.58	7.98	10.95
adjCRV	-0.032	-0.30	-0.25	0.10	0.15
CRR	-0.009	-0.25	-0.20	0.14	0.17

Table 12: (cont.) Empirical results using the delta-normal method. Ex ante length = 500 days, confidence level = 95%

	$IQR_{0.95\text{-}0.05}$	$IQR_{0.975\text{-}0.025}$	t-stat	$H_0^{95\%}$	$H_0^{90\%}$
CRV	12.6	16.5	0.255	acc	acc
adjCRV	0.35	0.45	-0.242	acc	acc
CRR	0.34	0.42	-0.083	acc	acc

Table 13: Empirical results using the historical simulation method. Ex ante length = 500 days, confidence level = 95%

	mean	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.95}$	$Q_{0.975}$
CRV	1.98	-6.78	-5.28	8.93	11.32
adjCRV	-0.041	-0.30	-0.26	0.11	0.16
CRR	-0.035	-0.30	-0.23	0.11	0.14

Table 14: (cont.) Empirical results using the historical simulation method. Ex ante length = 500 days, confidence level = 95%

	$IQR_{0.95\text{-}0.05}$	$IQR_{0.975\text{-}0.025}$	t-stat	$H_0^{95\%}$	$H_0^{90\%}$
CRV	14.2	18.1	0.329	acc	acc
adjCRV	0.37	0.46	-0.322	acc	acc
CRR	0.34	0.45	-0.295	acc	acc

Table 15: Empirical results using the delta-normal method. Ex ante length = 500 days, confidence level = 99%

	mean	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.95}$	$Q_{0.975}$
CRV	0.59	-9.85	-8.67	10.74	15.20
adjCRV	-0.050	-0.71	-0.54	0.30	0.32
CRR	-0.029	-0.90	-0.44	0.31	0.37

Table 16: (cont.) Empirical results using the delta-normal method. Ex ante length = 500 days, confidence level = 99%

	IQR 0.95-0.05	IQR 0.975-0.025	t-stat	$H_0^{95\%}$	$H_0^{90\%}$
CRV	19.4	25.1	0.091	acc	acc
adjCRV	0.84	1.03	-0.125	acc	acc
CRR	0.74	1.27	-0.068	acc	acc

Table 17: Empirical results using the historical simulation method. Ex ante length = 500 days, confidence level = 99%

	mean	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.95}$	$Q_{0.975}$
CRV	6.80	-7.76	-6.55	24.28	26.45
adjCRV	-0.32	-2.06	-1.66	0.15	0.21
CRR	-0.29	-1.81	-1.52	0.17	0.24

Table 18: (cont.) Empirical results using the historical simulation method. Ex ante length = 500 days, confidence level = 99%

	IQR 0.95-0.05	IQR 0.975-0.025	t-stat	$H_0^{95\%}$	$H_0^{90\%}$
CRV	30.8	34.2	0.733	acc	acc
adjCRV	1.81	2.27	-0.475	acc	acc
CRR	1.69	2.05	-0.419	acc	acc

Table 19: Empirical results using the delta-normal method. Ex ante length = 250 days, confidence level = 95%.

	mean	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.95}$	$Q_{0.975}$
CRV	3.98	-3.82	-1.64	10.25	13.60
adjCRV	-0.10	-0.43	-0.33	0.026	0.041
CRR	-0.07	-0.30	-0.26	0.067	0.093

Table 20: (cont.) Empirical results using the delta-normal method. Ex ante length = 250 days, confidence level = 95%.

	IQR 0.95-0.05	IQR 0.975-0.025	t-stat	$H_0^{95\%}$	$H_0^{90\%}$
CRV	11.9	17.4	0.757	acc	acc
adjCRV	0.36	0.47	-0.678	acc	acc
CRR	0.32	0.39	-0.43	acc	acc

Table 21: Empirical results using the historical simulation method. Ex ante length = 250 days, confidence level = 95%.

	mean	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.95}$	$Q_{0.975}$
CRV	5.41	-0.17	0.39	10.78	12.43
adjCRV	-0.13	-0.44	-0.38	-0.005	0.001
CRR	-0.11	-0.42	-0.30	0.002	0.009

Table 22: (cont.) Empirical results using the historical simulation method. Ex ante length = 250 days, confidence level = 95%.

	IQR 0.95-0.05	IQR 0.975-0.025	t-stat	$H_0^{95\%}$	$H_0^{90\%}$
CRV	10.4	12.6	1.652	acc	**rej**
adjCRV	0.38	0.44	-1.177	acc	acc
CRR	0.30	0.43	-0.918	acc	acc

Table 23: Empirical results using the delta-normal method. Ex ante length = 250 days, confidence level = 99%

	mean	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.95}$	$Q_{0.975}$
CRV	4.59	-4.51	-3.75	20.15	21.29
adjCRV	-0.18	-1.18	-0.85	0.086	0.11
CRR	-0.13	-1.15	-0.71	0.16	0.17

Table 24: (cont.) Empirical results using the delta-normal method. Ex ante length = 250 days, confidence level = 99%

	IQR 0.95-0.05	IQR 0.975-0.025	t-stat	$H_0^{95\%}$	$H_0^{90\%}$
CRV	23.9	25.8	0.652	acc	acc
adjCRV	0.94	1.29	-0.444	acc	acc
CRR	0.87	1.32	-0.297	acc	acc

Table 25: Empirical results using the historical simulation method. Ex ante length = 250, confidence level = 99%

	mean	$Q_{0.025}$	$Q_{0.05}$	$Q_{0.95}$	$Q_{0.975}$
CRV	13.15	3.81	4.25	28.16	31.31
hline adjCRV	-0.52	-2.11	-1.73	-0.052	-0.026
CRR	-0.43	-1.78	-1.30	-0.031	-0.004

Table 26: (cont.) Empirical results using the historical simulation method. Ex ante length = 250, confidence level = 99%

	IQR 0.95-0.05	IQR 0.975-0.025	t-stat	$H_0^{95\%}$	$H_0^{90\%}$
CRV	23.9	27.5	1.800	acc	**rej**
hline adjCRV	1.67	2.09	-0.842	acc	acc
CRR	1.27	1.77	-0.725	acc	acc

B Empirical Findings - Distribution of Empirical Measures

In this appendix, we present the numerical cumulative distribution functions (CDFs) of the different empirical values used for the evaluation of CVaR.

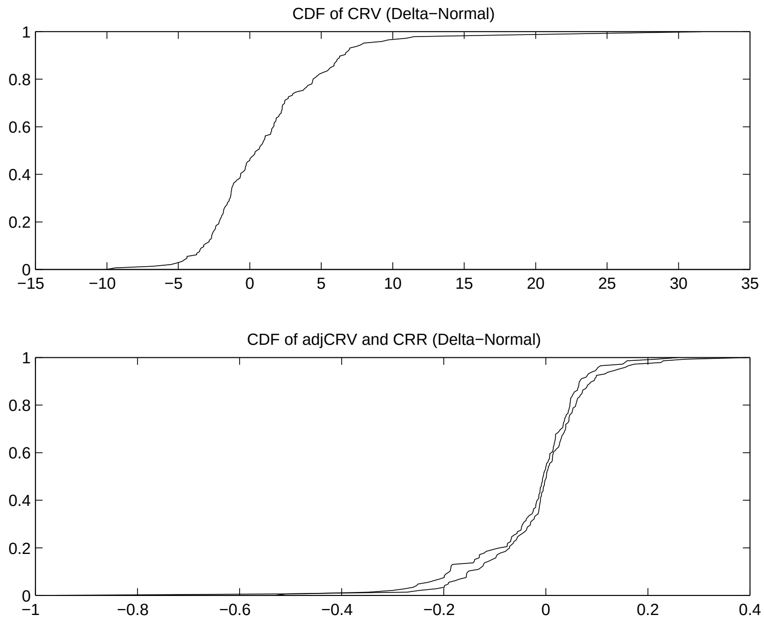

Figure 5: CDF of CRV, adjCRV and CRR using the delta-normal method. Ex ante length = 500 days, confidence level = 95%.

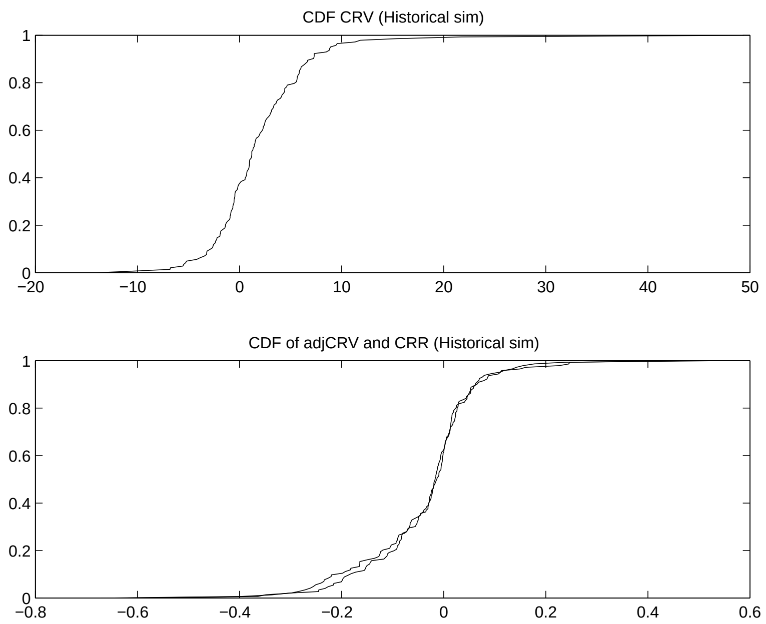

Figure 6: CDF of CRV, adjCRV and CRR using the historical simulation method. Ex ante length = 500 days, confidence level = 95%.

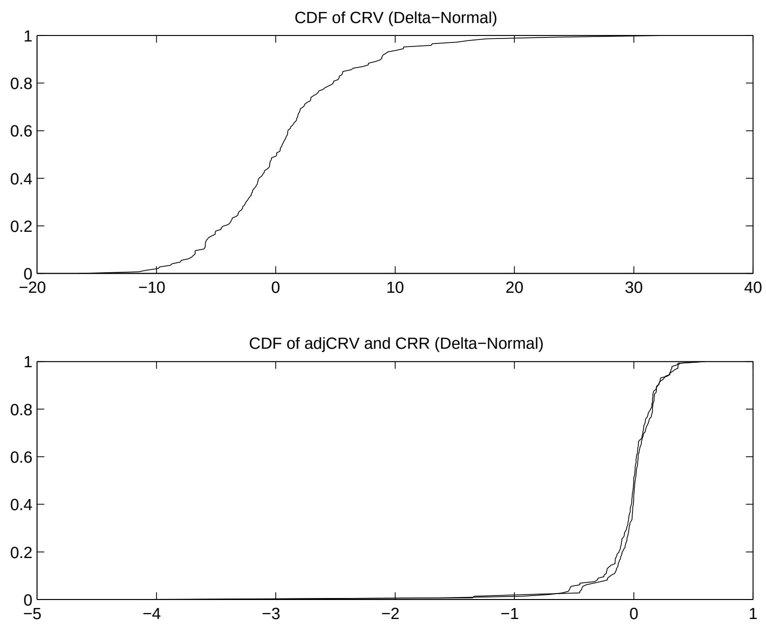

Figure 7: CDF of CRV, adjCRV and CRR using the delta-normal method. Ex ante length = 500 days, confidence level = 99%.

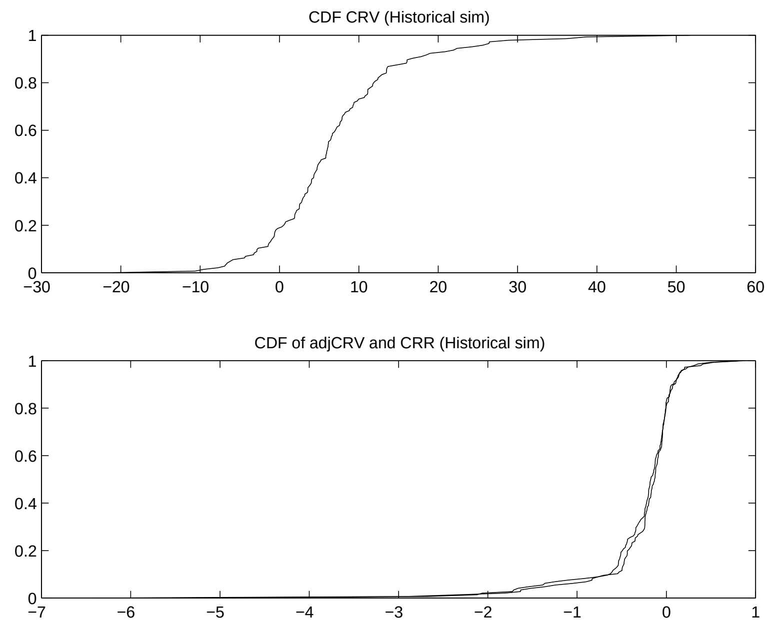

Figure 8: CDF of CRV, adjCRV and CRR using the historical simulation method. Ex ante length = 500 days, confidence level = 99%.

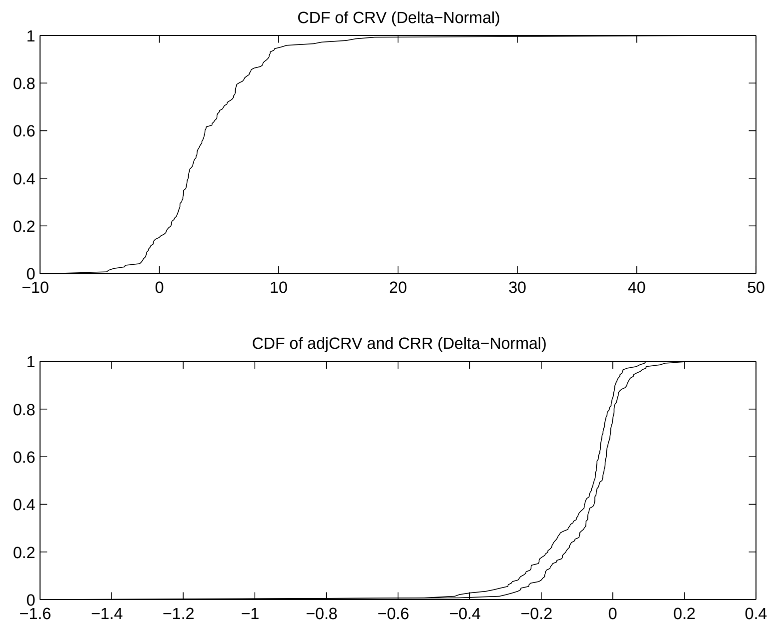

Figure 9: CDF of CRV, adjCRV and CRR using the delta-normal method. Ex ante length = 250 days, confidence level = 95%.

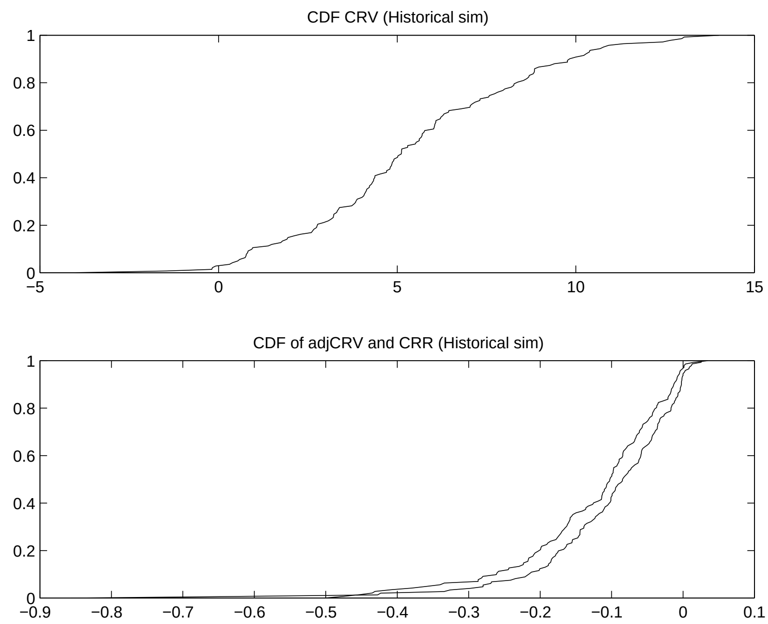

Figure 10: CDF of CRV, adjCRV and CRR using the historical simulation method. Ex ante length = 250 days, confidence level = 95%.

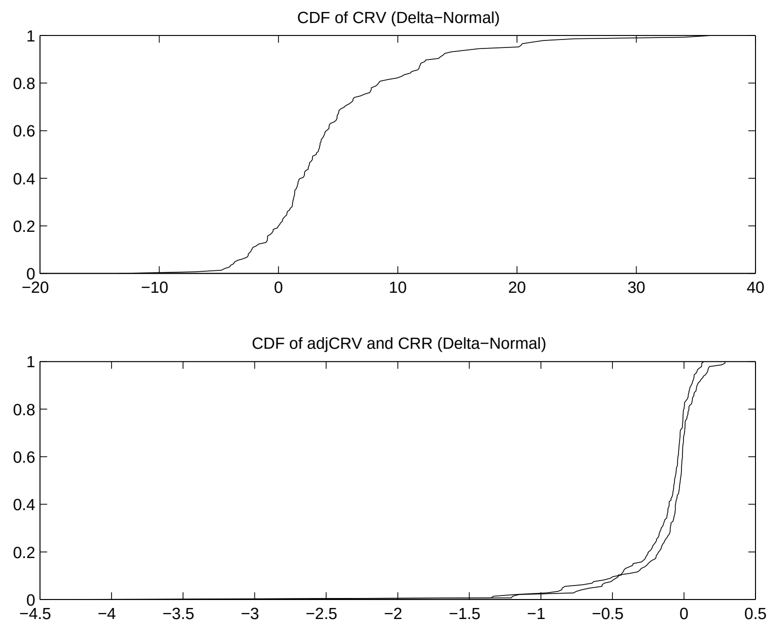

Figure 11: CDF of CRV, adjCRV and CRR using the delta-normal method. Ex ante length = 250 days, confidence level = 99%.

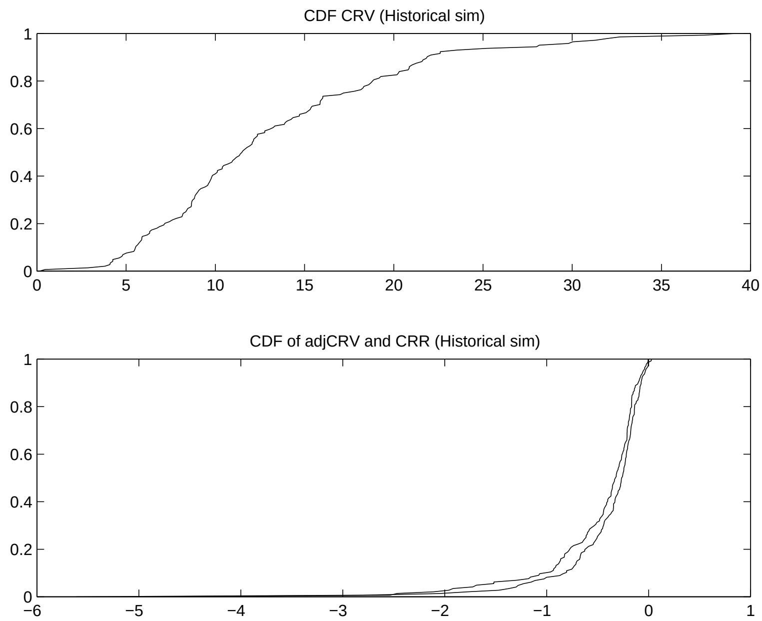

Figure 12: CDF of CRV, adjCRV and CRR using the historical simulation method. Ex ante length = 250 days, confidence level = 99%.

C Conditional Distributions

In Figures 13 - 22 the conditional return given that the return is less than VaR is plotted. The solid vertical line represents the VaR and the dashed vertical line represents the CVaR. The abscissa is scaled so that VaR is always located at 1 and CVaR at 1.5. Each plot shows the distribution of the returns over all assets and all asset classes considered in the study in the cases where the return is less than VaR. In order to be able to plot the returns for all assets and all points in time, each conditional return (or, to be exact, $r(t, \alpha) - \mathrm{CVaR}(t, \alpha)$, where $r(t, \alpha)$ is the conditional return at time t for asset α) is normalized with respect to the variable $\kappa(t, \alpha) = \mathrm{VaR}(t, \alpha) - \mathrm{CVaR}(t, \alpha)$. This makes returns over all assets and over time directly comparable.

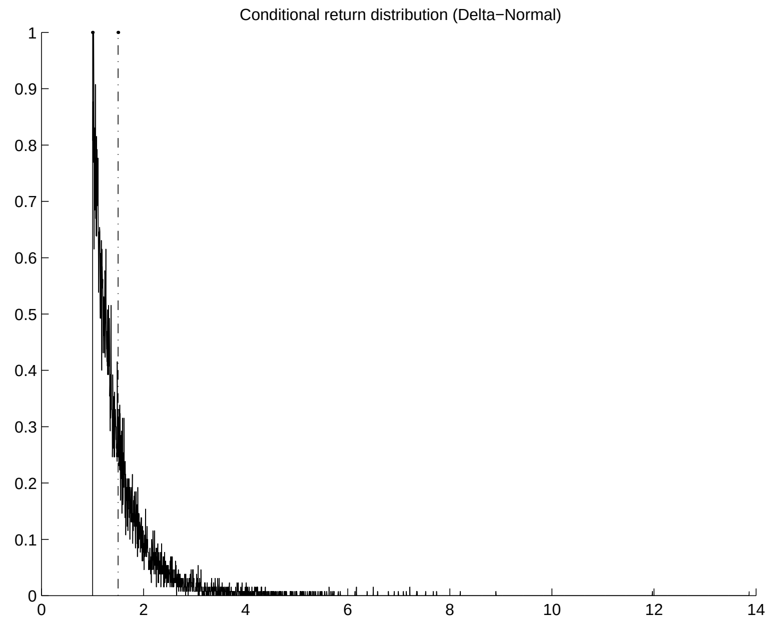

Figure 13: Conditional returns distribution using the delta-normal method. Ex ante length = 500 days, confidence level = 95%.

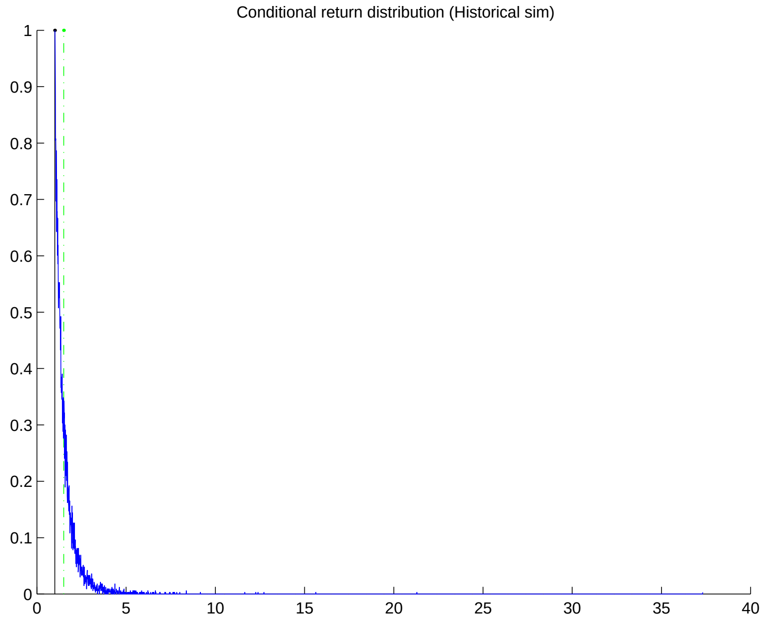

Figure 14: Conditional returns distribution using the historical simulation method. Ex ante length = 500 days, confidence level = 95%.

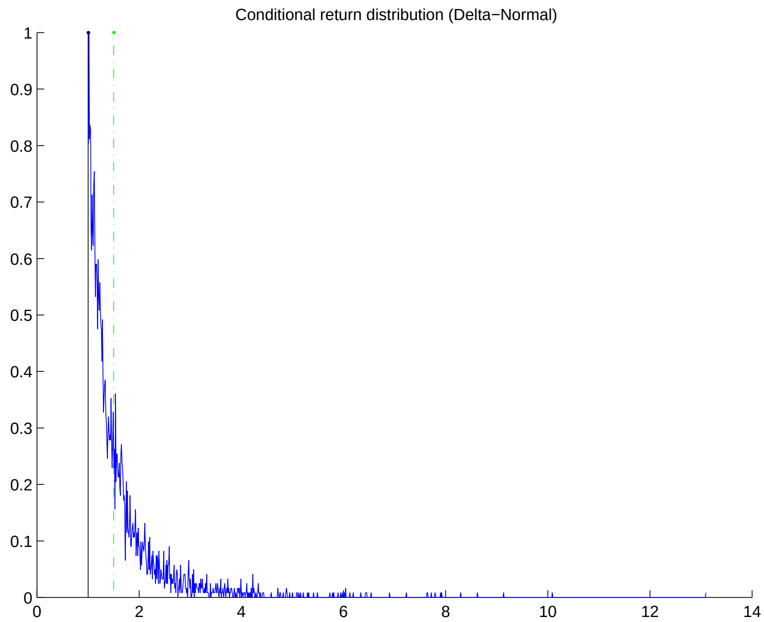

Figure 15: Conditional returns distribution using the delta-normal method. Ex ante length = 500 days, confidence level = 99%.

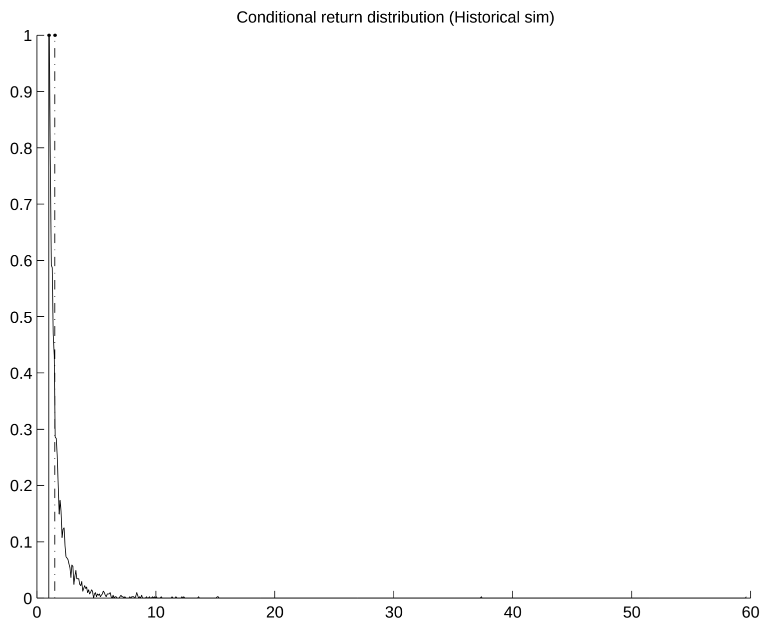

Figure 16: Conditional returns distribution using the historical simulation method. Ex ante length = 500 days, confidence level = 99%.

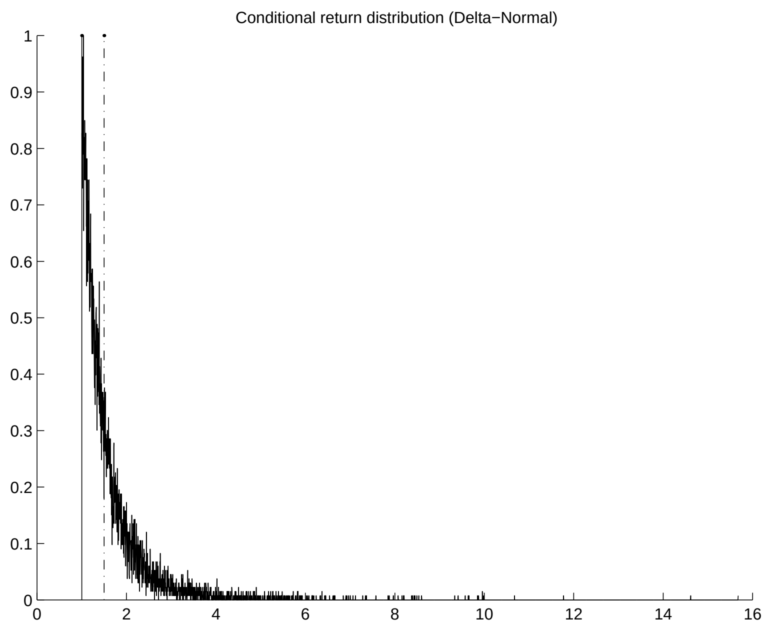

Figure 17: Conditional returns distribution using the delta-normal method. Ex ante length = 250 days, confidence level = 95%.

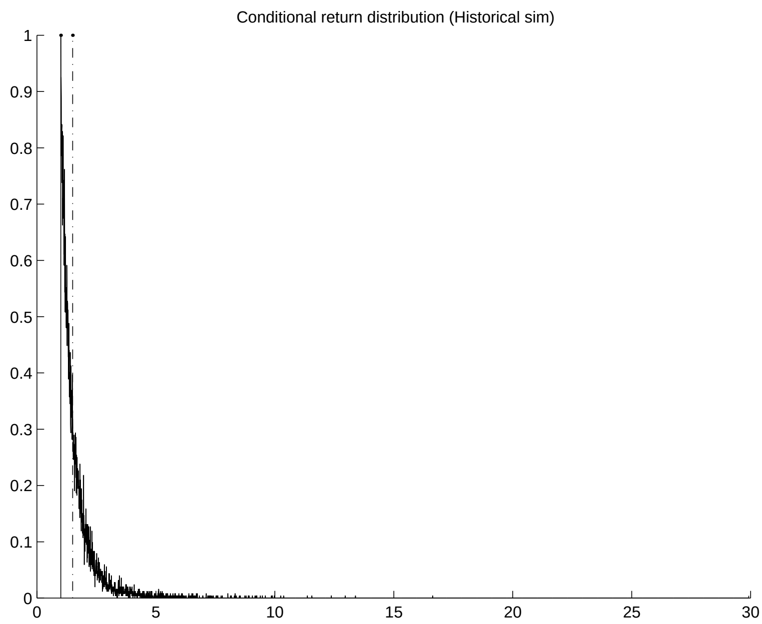

Figure 18: Conditional returns distribution using the historical simulation method. Ex ante length = 250 days, confidence level = 95%.

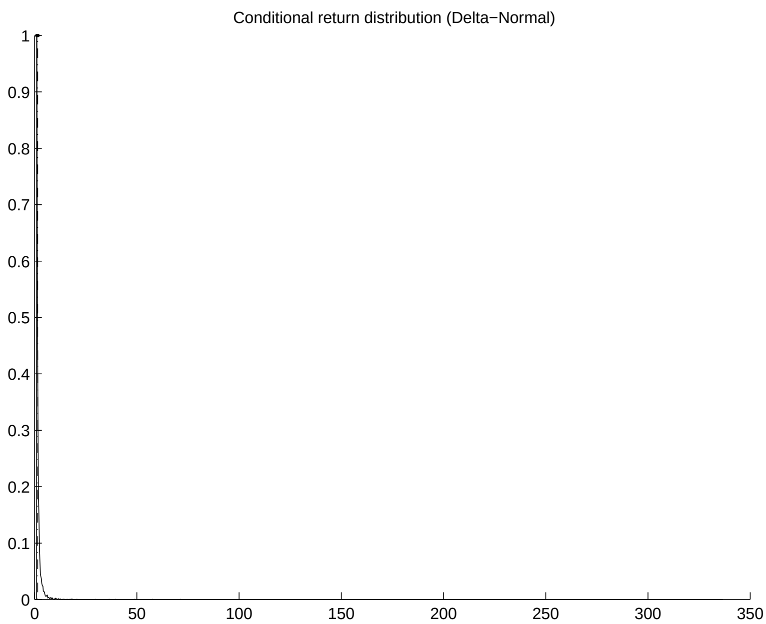

Figure 19: Conditional returns distribution using the delta-normal method. Ex ante length = 250 days, confidence level = 99%.

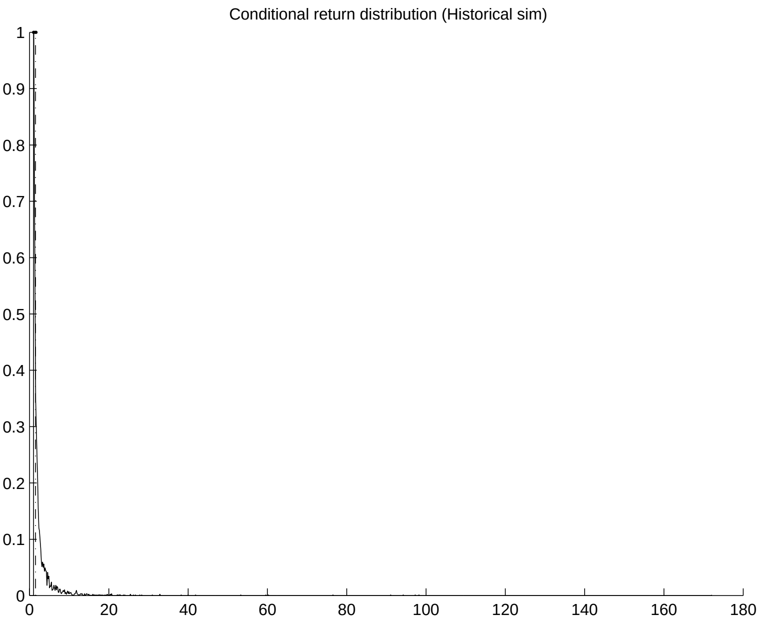

Figure 20: Conditional returns distribution using the historical simulation method. Ex ante length = 250 days, confidence level = 99%.

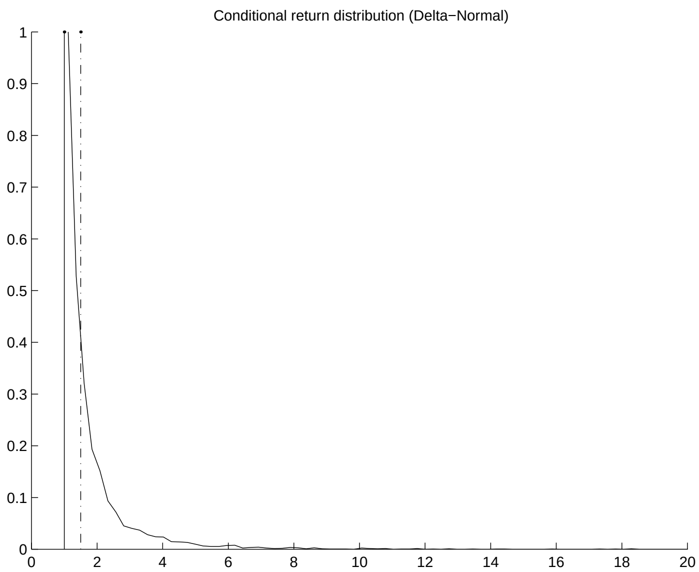

Figure 21: Conditional returns distribution using the delta-normal method. Ex ante length = 250 days, confidence level = 99%. Zoomed. Eight extreme returns are left out due to the zoom.

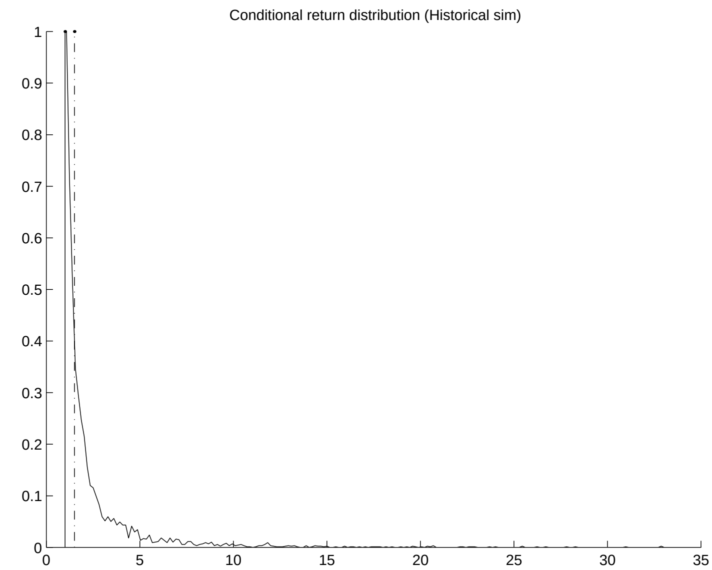

Figure 22: Conditional returns distribution using the historical simulation method. Ex ante length = 250 days, confidence level = 99%. Zoomed. Twelve extreme returns are left out due to the zoom.